Computational Modeling of Multimodal I/O in Simulated Cockpits

Final Report, Project N00014-96-1-0467

David E. Kieras
University of Michigan

EPIC Report No. 14 (TR-01/ONR-EPIC-14)

May 30, 2001

20010611 025

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE May 30, 2001	3. REPORT TYPE AND DATES COVERED Final June 1, 1996–Dec 30, 2000

4. TITLE AND SUBTITLE
Computational Modeling of Multimodal I/O in Simulated Cockpits

5. FUNDING NUMBERS
N000014-96-1-0467

6. AUTHOR(S)
David E. Kieras

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Michigan
Division of Research Development and Administration
Ann Arbor, MI 48109-1274

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-01/ONR-EPIC-14

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release: Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes the results of a project on modeling the effects of localized 3-D sound to facilitate performance in a complex cockpit-like dual task. This task that had been previously observed to produce a significant automation deficit effect: when one of the tasks has to be resumed at short notice, the human operator takes some time to "catch up" and reach the normal steady-state level of performance in the task, apparently because it takes visual search and inspection to identify the proper object on the display to process. Providing a localized sound cue to identify the proper object alleviates the automation deficit effect to some extent. Constructing computational cognitive models that include representation of the perceptual-motor systems underlying performance showed that the benefit appears to be due to low-level orienting reflex eye movements rather than high-level strategic use of the sound information.

14. SUBJECT TERMS

15. NUMBER OF PAGES
10

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT

Computational Modeling of Multimodal I/O in Simulated Cockpits

David Kieras
Artificial Intelligence Laboratory
Electrical Engineering and Computer Science Department
University of Michigan

Abstract

This report summarizes the results of a project on modeling the effects of localized 3-D sound to facilitate performance in a complex cockpit-like dual task. This task that had been previously observed to produce a significant automation deficit effect: when one of the tasks has to be resumed at short notice, the human operator takes some time to "catch up" and reach the normal steady-state level of performance in the task, apparently because it takes visual search and inspection to identify the proper object on the display to process. Providing a localized sound cue to identify the proper object alleviates the automation deficit effect to some extent. Constructing computational cognitive models that include representation of the perceptual-motor systems underlying performance showed that the benefit appears to be due to low-level orienting reflex eye movements rather than high-level strategic use of the sound information

Project Purpose and Goals

Background

This project was a collaboration with James A. Ballas at the Naval Research Laboratory (NRL). The basic purpose of the project is to follow up on some leads discovered in the Kieras & Meyer ONR project on the EPIC computational theory for dual-task performance (N00014-92-J-1173) conducted at the University of Michigan (UM). Kieras & Meyer constructed and tested models for the experiments conducted by Ballas and his coworkers (Ballas, Heitmeyer, & Perez, 1992a, b) on a simulated cockpit task. The task involves a tracking task on one display and a tactical decision-making task of classifying targets on a separate radar-like display; the tactical decision-making task is sometimes automated, and sometimes must be done by the human operator.

Ballas had observed an *automation deficit* effect, a temporary depression of performance when the operator must resume manual control of the formerly automated tactical decision task task with short notice. In an informal collaboration with Ballas, Kieras & Meyer were able to model this deficit, and explain it as a result of several limitations on human performance, but with the need for eye movements between the two displays as being a major contributor. That is, a visual event on the tactical task display would occur (e.g. a "blip" would change color), and the operator must then do visual search to pick a relevant object for inspection and response.

Can localized sound reduce automation deficit?

The focus of the project work was on the use of synthetic 3-D sound, delivered to the operator over headphones, that would provide a localized sound cue that designated which display object should be looked at. Since this cue could direct visual attention directly, it appeared that localized sound could be effective in reducing the visual search and extraneous eye movements, and thus reduce the automation deficit.

The NRL part of the project was to collect new empirical data on the automation deficit phenomenon and how it might be affected by the use of multiple sensory modalities, especially advanced auditory displays, that might reduce the need for eye movements. The UM part of the project was to advise on the empirical program and construct new models that account for the results in the general context of the EPIC cognitive architecture. The joint responsibility was to arrive at new and advanced understanding of how multimodal displays might help reduce pilot workload. In the course of this project, further work was done on extending the EPIC architecture and applying it to modeling dual tasks; this part of the work was jointly supported under other ONR funding, Grant Number N00014-92-J-1173.

Problems encountered

More so than most research projects, this project did not go according to plan. A basic assumption of the project was that there would be a tight iterative loop between the modeling program at UM and the empirical experimentation and data collection at NRL. However, after the first preliminary experiments, the empirical program at NRL ran into serious delays, especially extended delays over normally-routine human subjects approval. As a result, the key studies in the project were seriously delayed and there was not time to collect a complete and problem-free set of data on auditory localization. In the meantime, a round of modeling work on auditory localization was done, but the results were not put into final form pending a more complete set of data. These extended delays led to delays in the UM part of the project, which were handled by project extensions and budget reductions. When it became clear that more complete auditory localization data could not be collected in the available time, a second round of modeling work on auditory localization was completed, and useful results were obtained and distributed in the form of a technical report. This work, for the first time, ties a body of literature on auditory localization directly into a modern computational cognitive architecture.

Accomplishments

More background on the task

Some additional explanation of the task is needed to provide background for the rest of this report. The task was developed by Ballas, Heitmeyer, & Perez (1992a, b) to resemble a class of multiple tasks performed in combat aircraft in which the subject must both perform a task such as tracking a target, and at the same time keep up with the tactical situation using sensors such as radar, with partial automation support by an on-board computer. Figure 1 shows a sketch of the display. The right hand box contains a pursuit tracking task in which the circle cursor must be kept on the target with a joystick operated with the right hand. The left-hand box is a radar-like display that contains a tactical decision task in which objects (called *tracks*) must be classified as hostile or neutral based on their behavior, and the results entered by means of a keypad under the left hand. These objects appear as icons that represent fighter aircraft, cargo airplanes, and SAM sites. A number identifies each object on the display.
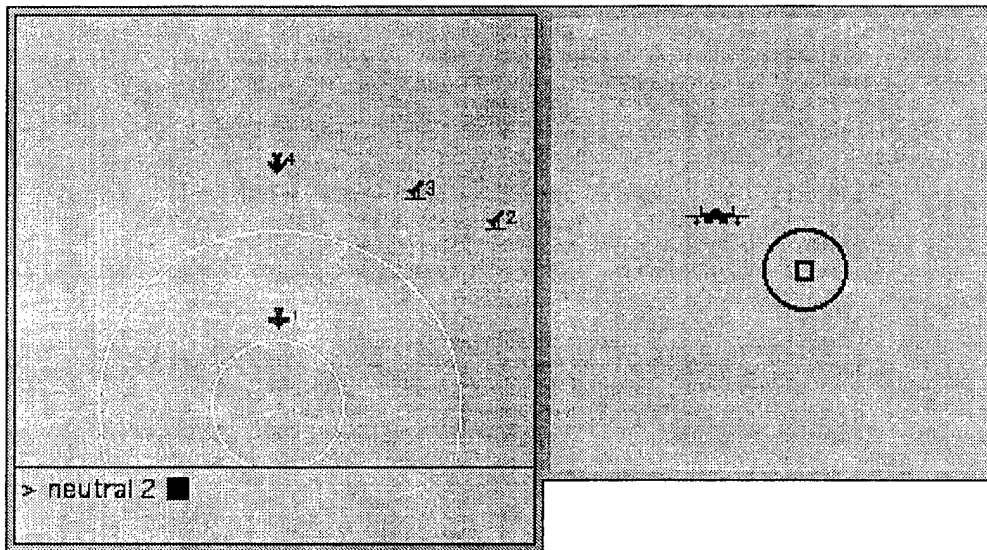
Figure 1. Screen shot of display, showing feedback from input keystrokes.

The blips appear near the top of the display, and then move down. The fictitious on-board computer attempts to classify each blip, indicating the outcome after some time by changing the blip color from black to red, blue, or amber. These color changes are termed *events* because these color changes are the stimuli to which the subject must respond. If the blip changes to red (hostile) or blue (neutral), the subject must simply confirm the computer's classification by typing a code key for the hostile/neutral designation followed by the key for the blip number. If the blip changes to amber, the subject must observe the behavior of the blip and classify it based on a set of rules, and then type the hostility designation and blip number. After the response, the blip changes color to white, and then disappears from the display 10 sec later. The basic dependent variable is the reaction time to the events, measured from when a blip changes color to when each of the two keystrokes are made in response.

From time to time during the task, the tracking task would become difficult, and the on-board computer would take over the tactical task, signaling when it did so. The computer would then generate the correct responses to each blip at the appropriate time, with the color changes showing on the display as in the manual version of the task. Later, the tracking would become easy again, and the computer would signal with a loud buzzer sound and then return the tactical task to the subject to perform. How subjects dealt with the transition was measured by recording the time required to respond to the individual events, counting from when they had to resume the tactical task.

In the scenarios resulting in automation deficit effects, the blips and color-change events within an epoch were not uniformly spaced in time; rather they occurred in two waves, the first when the task had to be resumed in manual mode at the beginning of each epoch, and the second about two-thirds of the way through the epoch. Some of the event time structure was fixed, with the remaining allowed to vary stochastically within the overall two-wave structure.

The automation deficit effect. Ballas et al. (1992a,b) observed an automation deficit effect, in which during the resumption panic phase, the period after resuming the tactical task, subjects produced longer response times for matched events compared to their normal steady-state manual performance, that is, the events at the clump panic at the second wave. Thus, as shown in
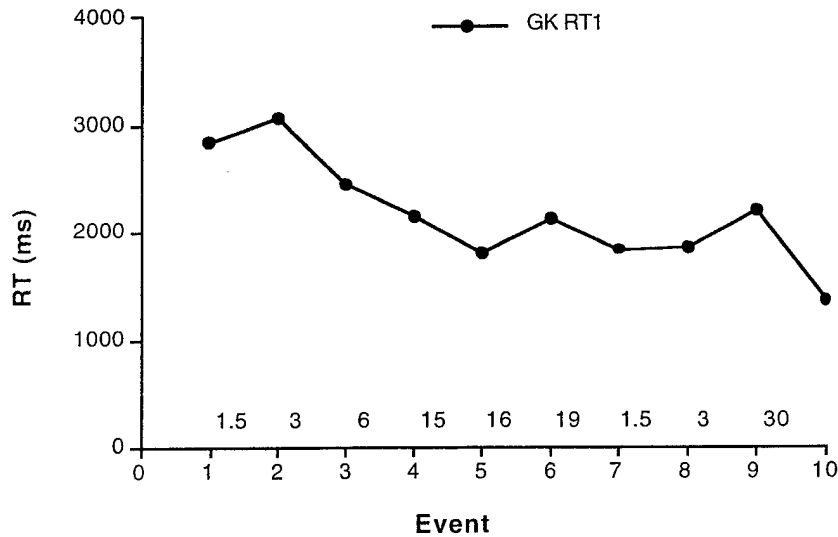
3

Figure 2. Illustration of the automation deficit effect, using data from Ballas et al. (1992a, b). First response times (RT1) are shown for each event (blip color change) after the tactical task is resumed. The numbers above the x-axis are the interval in seconds between each pair of events. The event sequence 1, 2, and 3 have the same inter-event spacing as event sequence 7, 8 and 9. The automation deficit effect appears as elevated RTs for the first events (e.g. 1) relative to the comparison events (e.g. 7) after the task has been underway for some time.

Figure 2, the times for Events 1, 2, and 3 are longer than the matched Events 7, 8, and 9, producing an overall descending shape to the RT profile. The reaction times for the first few events during the resumption panic and catching-up phase are substantially longer than those for later events of similar structure during the clump panic and its catching-up phase. Since Ballas et al. had arranged for Events 1 and 7 to be exactly matched in terms of the type of blip, they reported the automation deficit effect in terms of simply the difference between the RT for Event 1 and Event 7, which is 1312 ms for the first response keystroke. This effect represents some of the serious concerns about possible negative effects of automation in combat situations; if the automation fails, the operator can lack situation awareness, and it might take a long time to "catch up."

Verifying the automation-deficit model

Some of the first work in the project was verifying some of the basic assumptions and mechanisms in the EPIC models for the task. The automation deficit effect itself is explained as due to a "catching-up" process: when the signal to resume the task appears, the human must resume monitoring the previously-ignored display, and then attempt to determine which track to move the eyes to and start processing. Because the display was not monitored in the meantime, the first choice might be incorrect or not the highest-priority track. This extra time to get to the highest-priority track is the elevated reaction time, the automation deficit. The size of the deficit should depend on the number of tracks that the human would have to examine and process at the time of resumption. Thus, the greater the event density, or workload, at the time of resumption, the greater the automation deficit. Thus, to test the model, Ballas's group collected data in the task using a scenario in which the workload was very low at the time of task resumption - the resumption signal occurred, and some time later, only one track required processing. The result was that there was no automation deficit - the first track after resumption was processed as quickly as later tracks. This result was predicted by the previously developed EPIC models, simply by driving them with the

4

low-workload scenario. These results confirmed the basic explanation of the automation deficit. These results are presented in Kieras, Ballas, & Meyer (2001).

A second body of confirmation work addressed an assumption of the EPIC models that the tracking task was suspended while the tactical task was performed. Although in modeling other cases of dual-task performance we have found it important to interleave the two tasks, so that they were performed concurrently, the Ballas task is complex enough that we thought it more reasonable to assume that tracking was suspended. This assumption was supported by work done at NRL showing that the frequency of tracking movements was substantially reduced in the vicinity of tactical task responses. These results appeared in Ballas, Kieras, Meyer, Stroup, & Brock (1999).

Development of the bracketing heuristic

While the original work on auditory multimodal interfaces was delayed, this project stimulated some additional work on the application of computational cognitive modeling to practical problems. These concepts are as important as the originally planned work on multimodal interfaces, if not more so. A brief explanation can be given here. More complete discussion can be found in Kieras & Meyer (2000), and an example of their use appears .

A major accomplishment of current cognitive psychology is several computational cognitive architectures. These are software packages that represent the mechanisms underlying human cognition and perceptual-motor performance. A simulation model of a human performing a task can be constructed by "programming" the architecture to do the task. Running the simulated human on the task produces a stream of simulated behavior which can be compared to existing data to test the accuracy of the simulation, or used to predict what human performance would be like if the task does not exist yet. For example, if a human-machine system is being designed, such a simulation model could be used to predict how well humans would performing in using the system; the results could be used to refine the design of the system functionality or human interface to produce better performance.

However, there is a significant barrier to the application of computational models for human performance to system design, and this is what we have called the strategy identification problem. In order to perform a task, the human must follow a task strategy, or procedure, specific to the task and the design of the system and its interface. In order to construct a model of the human using the system, we have to have a description of the task strategy to use to program the cognitive architecture. Our work with the EPIC architecture had made it clear that in many high-performance tasks, even very simple ones, the human operator can apply a wide range of task strategies to optimize task performance over many different criteria. We discovered that the choice of strategy would make a huge difference in the level of predicted performance. In working with observed data on human performance, we could identify many features of the humans' task strategies by comparing details of their performance with the performance of models using different candidate strategies, and thus determine which strategy the humans appeared to be following. Clearly, the degree of practice or training and the motivation of the humans are the major determinants of how enterprising their strategies are.

The problem is that if one is attempting to predict the performance of a human-machine system in advance, one has to use a strategy in the model, but there is no observed data to be used to pick the strategy. Given that the choice of strategy makes a big difference in predicted performance, how can we make useful predictions given the lack of knowledge of the detailed strategy users will actually learn or apply to the future system?

5

Our answer was to propose a bracketing heuristic: We can build two models: a fastest-possible model that uses a task strategy that drives the architecture at the highest performance level possible, and a slowest-reasonable model that directly reflects the task structure and nominal requirements without any "bells and whistles" that maximize performance. The observed or actual performance would thus be "bracketed" by the fastest and slowest models; rather than try to predict exact performance by dubious guesses of future task strategy, we could confidently predict the upper and lower bounds of performance with a systematic approach based only on the cognitive architecture and the task structure and requirements.

If the predicted performance was satisfactory with both models, then one could be confident that the actual system would perform acceptably with even routine training and motivation of the users. If neither model would be fast enough, then the system design was at fault, and would have to be corrected before any degree of training or motivation would suffice. If the bracketing models bracket the desired level of performance, then more work is needed to either improve the design, or to determine if the training and motivation factors will enable users to dependably work at a higher level of performance.

This project provided a valuable stimulus and exercise-ground for this concept. Bracketing models were constructed for the original Ballas task data, but then the same models were applied to brand-new data collected on the same task but with different scenarios and instructional conditions. The bracketing predictions were generated prior to inspection of the data. In one case, the scenario was expected to eliminate the automation defect by reducing the event density at the time of task resumption. The bracketing models showed exactly this effect. A second task demonstrated automation deficit effects under somewhat different conditions and task scenario; again the a-priori bracketing models successfully bracketed the observed data. The bracketing approach was then applied to account for the effects of the localized sound cuing. A similar exercise was performed on the original Ballas task data to compare the effects of two interfaces which differed in the extent to which they supported *direct manipulation* (Kieras, Meyer, & Ballas, 2001).

This was a successful demonstration of an additional idea: that bracketing could support scientific explanations of phenomena in the data, based on the cognitive architecture, without the difficult and time-consuming process of iteratively constructing and testing models that attempt to fit the data precisely. For example, the conclusion was that the effect of localized sound was an across-the-board speed up in task performance, and could be best accounted for by an auditorily-triggered eye movement to the sound source. This conclusion was supported by the results that only the bracketing models incorporating this assumption could produce this effect, and only if the eye was not already fixated on tactical task objects.

Using the bracketing approach was far faster and more economical of modeling effort than iteratively constructing models that directly fit the observed data. Note that conclusions drawn from fitted models would be more limited, in that all of the strategy assumptions in fitted models would qualify the conclusions about the use of localized sound. Thus the conclusion based on only the bracketing models are also stronger and more general conclusions.

Modeling localized auditory cuing

This work, for the first time, ties a body of literature on auditory localization directly into a modern computational cognitive architecture. The model analysis was conducted using the powerful bracketing heuristic described above. The work is presented in Kieras, Ballas, & Meyer (2001).

Ballas collected data in a version of the task in which when a track changed color, a localized sound would be emitted by the track; the sound also represented the type of track - e.g. a siren

6

sound for a fighter. Furthermore, the sound was emitted by the highest-priority (first color-changed) track, thereby disambiguating the visual display. This was a "kitchen-sink" approach intended to give the sound cue every possible chance to lead to greatly improved task performance. A facilitating effect of the sound cuing was found, as presented in Ballas, Brock, Stroup, Kieras, and Meyer (1999).

The first problem with modeling the effects of localized sound was that originally EPIC (like all other extant cognitive architectures) did not have any representation of localized sound sources; the internal working memory representation of objects arranged in space was purely visual in content. Adding this feature to EPIC required generalizing the internal spatial working memory system to include both auditory and visual information. The architecture could then be programmed to make eye and hand movements to objects based on either visual or auditory location information, but that if the more accurate visual location information were available, it would be used instead of auditory information.

The more difficult problem was then determining how the localized sound cue information could be used in the task. Many models for a variety of hypotheses could be built, iteratively fit to the data, and evaluated. However, the bracketing heuristic made it possible to conduct this part of the work rather rapidly once the EPIC architecture had been generalized to include localized sound, and the decision made to work with the available empirical data.

In summary, the available empirical data showed the presence of a small across-the-board advantage of the sound cuing. The hoped-for large-scale benefits of sound did not seem to be present. To determine what the effect of localized sound cuing would be, three sets of bracketing models were constructed. The first was a baseline model that did not use sound at all, but was compatible with the use of sound, and followed a task strategy based on the earlier work with the task. The second was a modified version of the baseline model that assumed that the sound cue was used in the strategic process of choosing which track to process - the sound cue identified both by location and sound quality which was the highest priority track to process. Relative to the no-sound model, this model predicted a drastically reduced automation deficit because it directed the human's attention directly to the proper track at the start of task resumption. However, the remaining track responses would not be facilitated at all, because the visual information normally available was adequate to do the task. Thus, this "sound-selection" model could not explained the faster across-the-board effect of sound cuing.

The third model required an addition to the EPIC architecture: a sound onset or change could trigger a reflex eye movement to the location of the sound. In the other models, a stimulus selection strategy would choose a track object to move the eye to for detailed examination. However, the sound-reflex mechanism would "automatically" move the eye to the proper track, without any cognitive-strategic decision-making. This one, simple, assumption was adequate to account for the both the uniform nature and relatively small size of the effect, being similar to results in the literature observed in tasks in which simple visual choice reaction tasks are facilitated by a localized sound cue to stimulus location. What is especially interesting is that the eye-movement trigger mechanism was added into the architecture without any detailed parameter-fitting attempts, using a-priori values only, and it proved to produce effects whose magnitude is approximately correct.

The original goal was to gain insight into how multimodal display systems in cockpits could be useful. These results provide such insights: Using localized sound might well be valuable, but will apparently yield its benefits from low-level orienting responses (e.g. reflexive eye movements). At least in this task, the use of the sound cue in higher-level decision making strategies was not demonstrated, and the model analysis shows that such involvement is not necessary to explain the effects.

Summary of Accomplishments

In conclusion, the contributions of this project can be summarized as follows:

- Further work was done to demonstrate the applicability to important military tasks of an important computational cognitive architecture, EPIC, that explicitly represents human perceptual-motor mechanisms.

- Some implications and assumptions of earlier EPIC models of a complex cockpit-like dual-task situation that explained to automation deficit effect, were confirmed by empirical data collection and analysis.

- A valuable technique for applying computational models of human cognitive and performance, the bracketing heuristic, was developed and demonstrated using the tasks, data, and phenomena in this project.

- Localized auditory information was incorporated into EPIC, as a generalization of the internal spatial representation of objects in the environment.

- The benefit of supplying localized auditory information in a complex dual task was analyzed and demonstrated to operate at a relatively low level of triggering eye movements, and this mechanism was incorporated into the EPIC computational cognitive architecture.

Publications and Presentations

Book Chapters

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. C. Schraagen, S. E. Chipman, & V. L. Shalin (Eds.), Cognitive task analysis. Mahwah, NJ: Lawrence Erlbaum, 2000.

Kieras, D. E., Meyer, D. E., Ballas, J. A., & Lauber, E. J. (2000). Modern computational perspectives on executive mental control: Where to from here? In S. Monsell & J. Driver (Eds.), Control of cognitive processes: Attention and performance XVIII (pp. 681-712). Cambridge, MA: M.I.T. Press.

Published Presentations

Ballas J. A., Kieras, D. E. & Meyer, D. E. (1996). Computational modeling of multimodal I/O in simulated cockpits. In S. P. Frysinger & G. Kramer (Eds.) In Proceedings of the Third International Conference on Auditory Display. Xerox Palo Alto Research Center, Palo Alto, CA, Nov. 4-6, 1996, pp 135-136.

Kieras, D.E. (1997). The Role of Cognitive Task Analysis in the Application of Predictive Models of Human Performance. Presented at the NATO-ONR Workshop on Cognitive Task Analysis. Washington, D.C. Oct. 30 - Nov. 1, 1997.

Kieras, D.E., Meyer, D.E., James A. Ballas, J.A., & Lauber, E. J. (1998). Modern computational perspectives on executive mental control: Where to from here? Paper presented at Attention & Performance XVIII Cumberland Lodge, Windsor Great Park, England, July 13-17.

Ballas, J.R., Kieras, D.E., Meyer, D.E., Stroup, J., & Brock, D. (1999) How is tracking affected by actions on another task? In Proceedings of the Tenth Aviation Psychology Symposium, Columbus, Ohio, May 3-6, 1999.

Ballas, J., Brock, D. Stroup, J. Kieras, D. and Meyer, D. (1999) Cueing of Display Objects by 3-D Audio to Reduce Automation Deficit. In Proceedings of the Fourth Annual Symposium and Exhibition on Situational Awareness in the Tactical Air Environment. The Naval Air WarfareCenter, Patuxent River, MD, June 8-9, 1999, pp 100-110.

Kieras, D., Meyer, D., & Ballas, J. (2001), Towards demystification of direct manipulation: Cognitive modeling charts the gulf of execution. Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems. New York, ACM. Pp. 128 – 135.

Technical Reports

Kieras, D. & Meyer, D.E. (1998). The role of cognitive task analysis in the application of predictive models of human performance. (EPIC Tech. Rep. No. 11 (TR-98/ONR-EPIC-11). Ann Arbor, University of Michigan, Electrical Engineering and Computer Science Department. March 5, 1998.

Kieras, D.E., Meyer, D.E., Ballas, J.A., Lauber, E.J. (1999). Modern computational perspectives on executive mental processes and cognitive control. Where to from here? (EPIC Tech. Rep. No. 12, TR-98/ONR-EPIC-12). Ann Arbor, University of Michigan, Electrical Engineering and Computer Science Department. August 1, 1999.

Kieras, D.E., Ballas, J.A., & Meyer, D.E. (2001). Computational models for the effects of localized sound cuing in a complex dual task. (EPIC Tech. Rep. No. 13, TR-01/ONR-EPIC-13). Ann Arbor, University of Michigan, Electrical Engineering and Computer Science Department. January 31, 2001.

Unpublished Presentations

Kieras, D.E. (2000, March). Modeling Human Performance for System Design Evaluation. Presented at the Cognitive Science Colloquium series, Georgia Institute of Technology, March 17, 2000.

Kieras, D.E. (1997, November). Predicting Human Performance by Modeling Procedural Knowledge. Presented at the ONR Workshop on Cognitive Models in the Design and Operation of Complex Systems. November 13-15, 1997, Baltimore.

References

Ballas, J. A., Heitmeyer, C. L., & Perez, M. A. (1992a). *Direct manipulation and intermittent automation in advanced cockpits.* Technical Report NRL/FR/5534--92-9375. Naval Research Laboratory, Washington, D. C.

Ballas, J. A., Heitmeyer, C. L., & Perez, M. A. (1992b). *Evaluating two aspects of direct manipulation in advanced cockpits.* In Bauersfeld, P., Bennett, J., and Lynch, G., *CHI'92 Conference Proceedings: ACM Conference on Human Factors in Computing Systems,* Monterey, May 3-7, 1992.